AFRL-AFOSR-VA-TR-2016-0293

Reliable Function Approximation and Estimation

Rachel Ward
UNIVERSITY OF TEXAS AT AUSTIN
101 EAST 27TH STREET STE 4308
AUSTIN, TX 78712

08/16/2016
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/RTA2

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
12-08-2016	FINAL PROJECT REPORT	4-15-2013-4-14-2016

4. TITLE AND SUBTITLE
Reliable Function Approximation and Estimation

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-13-1-0125

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Rachel A Ward

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Rachel Ward
UNIVERSITY OF TEXAS AT AUSTIN 101 EAST 27TH
STREET STE 4308 AUSTIN, TX 78712

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
875 N. Randolph, Ste.325
Arlington Virginia, 22203
Email: info@afosr.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A - Approved for Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Exploiting the latent structure in many real-world signals can dramatically increase algorithmic robustness to both noise and missing data. The theory of compressed sensing shows that if a signal of interest is sparse --- well-approximated by some small subset of a dictionary of basis elements --- then the signal can be acquired from a reduced number of measurements and reconstructed using efficient convex programming techniques. However, the standard compressed sensing theory is valid only for a restrictive set of dictionaries, limiting the scope of applications. In this award, the PI developed a range of reliable and structure-aware sampling theorems based on the weighted sparsity model for real-world systems which are governed mostly by low-order interactions. The weighted sparsity model allows for more freedom than linear regression but provides sufficient structure to extend compressed sensing results to a wide class of infinite-dimensional problems. We discuss four key application domains for the methods developed in this project arising from this project: uncertainty quantification, image processing, matrix completion, and stochastic optimization.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U			

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Final report: Reliable function approximation and estimation

Rachel Ward, University of Texas at Austin, rward@math.utexas.edu

Exploiting the latent structure in many real-world signals can dramatically increase algorithmic robustness to both noise and missing data. The theory of compressed sensing shows that if a signal of interest is sparse — well-approximated by some small subset of a dictionary of basis elements — then the signal can be acquired from a reduced number of measurements and reconstructed using efficient convex programming techniques. However, the standard compressed sensing theory is valid only for a restrictive set of dictionaries, limiting the scope of applications.

During the tenure of this award, as anticipated, the PI developed a range of reliable and structure-aware sampling theorems based on the weighted sparsity model for real-world systems which are governed mostly by low-order interactions. The weighted sparsity model and weighted sampling allows for more freedom than linear regression but provides sufficient structure to extend compressed sensing results to a wide class of infinite-dimensional problems. We discuss four key findings arising from this project, as related to uncertainty quantification, image processing, matrix completion, and stochastic optimization, respectively.

In paper (P1), we consider the problem of function interpolation, and provide theoretical basics for weighted sparse approximation. We provided weighted stochastic sampling strategies for interpolating sparse or compressible expansions in orthogonal polynomial bases from a minimal number of pointwise function evaluations. Based on a model of weighted sparsity which we introduced, we provide error rates and choices of weights for regularization via weighted L1 minimization. We later extended this work to overcomplete dictionaries (P2) and refined the sample complexity analysis for Gaussian measurements in (P3). Our work has found interest and application in uncertainty quantification, namely the polynomial chaos approach, where one approximates the dependence of simulation model output on model parameters by expansion in an orthogonal polynomial basis.

In paper (P6) we considered the application of weighted sampling to medical imaging where one seeks to recover a good approximation of an images with sparsity in terms of its spatial finite differences and wavelet transform coefficients from a subset of measurements in the Fourier domain. We formulated the notion of local coherence in the discrete setting and, by bounding the inner product between Fourier and Haar wavelet basis elements in a certain way, provided near-optimal reconstruction guarantees with sampling frequencies from a fixed distribution where a frequency component is sampled with probability proportional to its squared magnitude and recovering an image via total variation minimization from such samples.

Matrix completion refers to the problem of recovering a low-rank matrix from a small subset of its elements, and we also applied the concept of weighted sampling to successfully extend the state-of-the-art results for matrix completion in papers (P4, P5). Matrix completion was previously known to be possible when the matrix satisfies a restrictive structural constraint—known as incoherence—on its row and column spaces. In these cases, the subset of elements is sampled uniformly at random. We showed that any rank-r matrix can be exactly recovered from as few as order O(n r) randomly chosen elements, provided this random choice is made according to a specific biased distribution: the probability of any element being sampled should be proportional to the sum of the leverage scores of the corresponding row and column.

Finally, we applied weighted sampling theorems to a seemingly very different application in large-scale machine learning: stochastic gradient descent. SGD is an iterative procedure for minimizing a high-dimensional function whereby at each step, one chooses an index and descends along the direction of the gradient of a constituent function, repeated until convergence to within a prescribed tolerance. In huge-scale optimization problems, stochastic gradient descent is an effective surrogate for full gradient descent, which is too expensive. The default sampling strategy in stochastic gradient methods is to sample component directions for descent uniformly at random. In reference (P7), we showed that re-weighting the sampling distribution so that components with larger variation are sampled with higher probability is necessary in order to improve convergence over uniform sampling, and obtain a linear dependence on average, as opposed to worst-case, smoothness among the constituent functions. Our results are based on a connection between SGD and the randomized Kaczmarz algorithm, which had until this point been studied essentially independently from SGD, allowed us to transfer ideas between the separate bodies of literature studying each of the two methods.

AFOSR Deliverables Submission Survey

Response ID:6659 Data

1.

1. Report Type

Final Report

Primary Contact E-mail
Contact email if there is a problem with the report.

rward@math.utexas.edu

Primary Contact Phone Number
Contact phone number if there is a problem with the report

5128005066

Organization / Institution name

University of Texas at Austin

Grant/Contract Title
The full title of the funded effort.

(YIP): Reliable function approximation and estimation

Grant/Contract Number
AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-13-1-0125

Principal Investigator Name
The full name of the principal investigator on the grant or contract.

Rachel A Ward

Program Manager
The AFOSR Program Manager currently assigned to the award

James Lawton

Reporting Period Start Date

04/15/2013

Reporting Period End Date

04/14/2016

Abstract

Exploiting the latent structure in many real-world signals can dramatically increase algorithmic robustness to both noise and missing data. The theory of compressed sensing shows that if a signal of interest is sparse --- well-approximated by some small subset of a dictionary of basis elements --- then the signal can be acquired from a reduced number of measurements and reconstructed using efficient convex programming techniques. However, the standard compressed sensing theory is valid only for a restrictive set of dictionaries, limiting the scope of applications.

During the tenure of this award, the PI developed a range of reliable and structure-aware sampling theorems based on the weighted sparsity model for real-world systems which are governed mostly by low-order interactions. The weighted sparsity model and weighted sampling allows for more freedom than linear regression but provides sufficient structure to extend compressed sensing results to a wide class of infinite-dimensional problems. We discuss four key application domains for the methods developed in this project arising from this project: uncertainty quantification, image processing, matrix completion, and stochastic optimization, respectively.

In paper (P1), we consider the problem of function interpolation, and provide theoretical basics for weighted sparse approximation. We provided weighted stochastic sampling strategies for interpolating sparse or compressible expansions in orthogonal polynomial bases from a minimal number of pointwise function evaluations. Based on a model of weighted sparsity which we introduced, we provide error rates and choices of weights for regularization via weighted L1 minimization. We later extended this work to overcomplete dictionaries (P2) and refined the sample complexity analysis for Gaussian measurements in (P3). Our work has found interest and application in uncertainty quantification, namely the polynomial chaos approach, where one approximates the dependence of simulation model output on model parameters by expansion in an orthogonal polynomial basis.

In paper (P6) we considered the application of weighted sampling to medical imaging where one seeks to recover a good approximation of an images with sparsity in terms of its spatial finite differences and wavelet transform coefficients from a subset of measurements in the Fourier domain. We formulated the notion of local coherence in the discrete setting and, by bounding the inner product between Fourier and Haar wavelet basis elements in a certain way, provided near-optimal reconstruction guarantees with sampling frequencies from a fixed distribution where a frequency component is sampled with probability proportional to its squared magnitude and recovering an image via total variation minimization from such samples.

Matrix completion refers to the problem of recovering a low-rank matrix from a small subset of its elements, and we also applied the concept of weighted sampling to successfully extend the state-of-the-art results for matrix completion in papers (P4, P5). Matrix completion was previously known to be possible when the matrix satisfies a restrictive structural constraint---known as incoherence---on its row and column spaces. In these cases, the subset of elements is sampled uniformly at random. We showed that any rank-r matrix can be exactly recovered from as few as order O(n r) randomly chosen elements, provided this random choice is made according to a specific biased distribution: the probability of any element being sampled should be proportional to the sum of the leverage scores of the corresponding row and column.

Finally, we applied weighted sampling theorems to a seemingly very different application in large-scale machine learning: stochastic gradient descent(SGD). SGD is an iterative procedure for minimizing a high-dimensional function whereby at each step, one chooses an index and descends along the direction of the gradient of a constituent function, repeated until convergence to within a prescribed tolerance. In huge-scale optimization problems, stochastic gradient descent is an effective surrogate for full gradient descent, which is too expensive. The default sampling strategy in stochastic gradient methods is to sample component directions for descent uniformly at random. In reference (P7), we showed that re-weighting the sampling distribution so that components with larger variation are sampled with higher probability is necessary in order to improve convergence over uniform sampling, and obtain a linear dependence on average, as opposed to worst-case, smoothness among the constituent functions. Our results are based on a connection between SGD and the randomized Kaczmarz algorithm, which had until this point been studied essentially independently from SGD, allowed us to transfer ideas between the separate bodies of literature studying each of the two methods.

Distribution Statement

This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

sf0298.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

final_report.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

(P*) Publications directly related to this proposal.
(O*) Publications indirectly related to this proposal.

(P1) Interpolation via weighted $\ell 1$ minimization. H. Rauhut and R. Ward. Applied and Computational Harmonic Analysis 40 (2), 2016. 321-351.

(P2) Compressive sensing with redundant dictionaries and structured measurements. F Krahmer, D Needell, and R Ward.
SIAM Journal on Mathematical Analysis 47 (6), 2015. 4606-4629.

(P3) The Sample Complexity of Weighted Sparse Approximation. B. Bah and R. Ward. IEEE Transactions on Signal Processing 64 (12), 2015. 3145-3155.

(P4) Completing any low-rank matrix, provably.
Y Chen, S Bhojanapalli, S Sanghavi, R Ward
Journal of Machine Learning Research 16, 2016. 2999-3034.

(P5) Coherent matrix completion. Y Chen, S Bhojanapalli, S Sanghavi, R Ward. ICML, 2014.

(P6) Stable and robust sampling strategies for compressive imaging. F Krahmer, R Ward. IEEE transactions on image processing 23 (2), 2014. 612-622.

(P7) Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm. D Needell, N Srebro, R Ward. NIPS, 2014.

(O1) One-bit compressive sensing with norm estimation. K Knudson, R Saab, and R Ward. IEEE Transactions on Information Theory 62 (5), 2016. 2748-2758.

(O2) An arithmetic–geometric mean inequality for products of three matrices. A Israel, F Krahmer, and R Ward. Linear Algebra and its Applications 488, 2016. 1-12.

(O3) Recovery guarantees for exemplar-based clustering. A Nellore and R Ward. Information and Computation 245, 2015. 165-180.

(O4) The local convexity of solving systems of quadratic equations. S. Sanghavi, C. White, and R. Ward. Results in Mathematics, 2016.

(O5) Relax, no need to round: Integrality of clustering formulations. P Awasthi, AS Bandeira, M Charikar, R Krishnaswamy, S Villar, R Ward. Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science.

(O6) A unified framework for linear dimensionality reduction in L1. F Krahmer and R Ward.
Results in Mathematics, 2014. 1-23.

(O7) Some deficiencies of χ2 and classical exact tests of significance. W Perkins, M Tygert, R Ward. Applied and Computational Harmonic Analysis, 2014.

(O8) Near-optimal compressed sensing guarantees for total variation minimization. D Needell and R Ward. IEEE transactions on image processing 22 (10), 2013. 3941-3949

(O9) Testing Hardy–Weinberg equilibrium with a simple root-mean-square statistic. R Ward and RJ Carroll. Biostatistics, 2013.

(O10) Stable image reconstruction using total variation minimization. D Needell and R Ward. SIAM Journal on Imaging Sciences 6 (2), 2013. 1035-1058.

2. New discoveries, inventions, or patent disclosures:

Do you have any discoveries, inventions, or patent disclosures to report for this period?

No

Please describe and include any notable dates

Do you plan to pursue a claim for personal or organizational intellectual property?

Changes in research objectives (if any):

The concepts of weighted sampling theorems for compressive sensing extended far beyond compressive sensing, and we had not realized the scope of the methods at the time. Thus, our research objectives branched out also to cover sampling strategies in stochastic optimization (P7) and matrix completion (P4 and P5).

Change in AFOSR Program Manager, if any:

At the beginning of my award, the program manager was Robert Bonneau. At the end, the program manager was James Lawton.

Extensions granted or milestones slipped, if any:

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Aug 12, 2016 08:11:45 Success: Email Sent to: rward@math.utexas.edu